

Statistics Limits Nonlocality

Avishy Y. Carmi and Daniel Moskovich

Center for Quantum Information Science and Technology &
Faculty of Engineering Sciences
Ben-Gurion University of the Negev, Beersheba 8410501, Israel.

ABSTRACT. A central theme in quantum mechanics is *nonlocality* which states that a pair of quantum systems which are shown not to be physically interacting may nevertheless be impossible to describe as independent entities. Quantum theory, however, forbids nonlocal correlations beyond a certain limit. Why is nature only so nonlocal and not more? Approaching the question from the direction of statistics and statistical inference, we identify a *statistical no-signaling principle* which states that no information may pass through a disconnected channel. We show this principle to be equivalent to the Tsirelson bound on nonlocality for the Bell–CHSH inequality.

Some of the predictions made by quantum mechanics appear to be at odds with common sense. Yet quantum mechanics remains the most precisely tested and successful quantitative theory of nature. It is therefore believed that even if quantum mechanics is someday replaced, any successor will have to inherit at least some of its “preposterous” but highly predictive principles. Perhaps the most counter-intuitive quantum mechanical principle is *nonlocality* [1]:

Nonlocality: A pair of quantum systems which are shown not to be physically interacting may nevertheless be impossible to describe as independent entities.

The mystery of nonlocality is not only to understand why nature is as nonlocal as it is, but also to understand why nature is not *more* nonlocal than it is. There are alternative *Non-Signaling* theories which permit nonlocality beyond the quantum limit [2, 3]; why doesn’t nature choose these theories over quantum mechanics? Several explanations have been proposed, but none is tight, *i.e.* none provides a necessary and sufficient condition for the quantum limit [4, 5]. We exhibit a protocol (an infinite oblivious transfer) which uses ‘superquantum NS-boxes’ to send messages through a disconnected channel, and we propose a principle which we call *statistical no-signaling* which states that such a communication is physically impossible. We show that statistical no-signaling in a bipartite setting is equivalent to *Tsirelson’s bound for the CHSH inequality* which we henceforth call *the quantum bound on nonlocality*. We thus provide a conceptual explanation for this bound. Our approach is different from others in that we use statistical techniques as opposed to probabilistic techniques—in particular we use Fisher information.

A famous application of nonlocality is to construct an 1-2 *oblivious transfer protocol* between two distant agents (A)lice and (B)ob. Alice and Bob each possess a mysterious box representing one half of the quantum system to be explained. Alice’s box might, for example, contain one half of a singlet state of spin- $\frac{1}{2}$ particles, with Bob’s box containing the other half [1, 6]. In addition, Alice possesses a pair of bits x_0 and x_1 , each of which is a zero or a one. Using boolean algebra and her boxes (the protocol will be described later), Alice encodes her pair of bits into a single bit $x^{(1)}$ which she sends across a classical channel to Bob. Bob wishes to recover either x_0 or x_1 , but Alice doesn’t know in advance which one. Bob uses the received bit $x^{(1)}$, his box, and some boolean algebra to construct an estimate y_i for his desired bit x_i . See Figure 1 later on.

What is the probability that Bob correctly estimates the bit he wished to know? He has two possible sources of knowledge—the bit $x^{(1)}$ he received from Alice, and some mysterious ‘nonlocal’ correlation between his box and Alice’s. The strength of such a nonlocal coordination between two systems is encapsulated by a number $c \in [-1, 1]$ called the *Bell-CHSH correlation* such that Bob’s probability of guessing

correctly is $(1 + |c|)/2$ (see Supplementary A). The *Bell–CHSH inequality* tells us that $|c| \leq 1/2$ classically [1, 6]. Mathematically, the statement of *nonlocality* is that c may violate the Bell–CHSH inequality. This has been supported by increasingly supported by experiments, culminating in a recent loophole-free verification [7]. We think of the Bell–CHSH correlation c as a measure of the *strength* of the nonlocality manifest in our boxes.

How large can c be? *Tsirelson’s bound* tells us that $|c|$ cannot exceed $1/\sqrt{2}$ in a world described by quantum mechanics [8]. This quantum bound on nonlocality:

$$(1) \quad |c| \leq \frac{1}{\sqrt{2}} ,$$

has been tested experimentally, with the current state of the art being an experiment by Kurtsiefer’s group which has achieved a value of c which is only 0.0008 ± 0.00082 distant from Tsirelson’s bound [5]. Such experimental evidence supports that Tsirelson’s bound indeed holds in the real world.

Tsirelson’s result is a specifically quantum mechanical fact for which there has been no good conceptual explanation. How fundamental is (1)? Must this inequality also hold for any future theory which might someday supersede quantum mechanics [9]? We are led to the following question:

Question: Can we find a plausible physical principle, independent of quantum mechanics, which is necessary and sufficient to guarantee that $|c| \leq 1/\sqrt{2}$?

The search for such a principle has a history of about 20 years. It was initially expected that the physical principle of relativistic causality (no-signaling) itself restricts the strength of nonlocality [10, 11, 12]. But then it was discovered that no-signaling theories may exist for which $|c| > 1/\sqrt{2}$. This led to the device-independent formalism of *No-Signaling (NS)–boxes* [2, 13] (see also [3]). In particular, maximum violation of the Bell–CHSH inequality is achieved by *Popescu–Rohrlich (PR)–boxes* which are consistent with Relativistic Causality. Why then, after all, does nature not permit (1) to be violated (as far as we know)? Several suggestions have been made. Superquantum correlations lead to violations of the Heisenberg uncertainty principle [14, 15], which is another seemingly purely quantum result. PR–boxes would allow distributed computation to be performed with only one bit of communication [16], which looks unlikely but doesn’t violate any known physical law. Similarly, in stronger-than-quantum nonlocal theories some computations exceed reasonable performance limits [17]. The principle of *information causality* [18] shows that no sensible measure of mutual information exists between pairs of systems in superquantum nonlocal theories. Finally, it was shown that superquantum nonlocality does not permit classical physics to emerge in the limit of infinitely many microscopic systems [19, 20]. Of these, only information causality and macroscopic locality give necessary conditions for the quantum bound and neither is known to be sufficient [4]. Thus these conditions do not single out quantum mechanics from amongst all possible nonlocal theories, as pointed out in [21].

We propose the following statement as a physical principle:

Statistical no-signaling: No information can pass through a channel whose output is independent of its input.

In this report we formulate a consequence of statistical no-signaling that is equivalent to (1), providing a sought-for conceptual explanation for the quantum bound on nonlocality. The novelty of our approach is our use of statistical methods.

Let $\mathbf{x} = \text{Bernoulli}(\theta)$ be a Bernoulli random variable held by Alice, which serves as our *information source*. We imagine $\theta \in [-1, 1]$ as encoding a message, perhaps in the digits of its binary expansion. Alice independently samples m values $\mathcal{A} \stackrel{\text{def}}{=} \{x_0, x_1, \dots, x_{m-1}\}$ from \mathbf{x} (the interesting case is $m \rightarrow \infty$) which she sends through a channel to Bob. Bob receives a set of values $\mathcal{B} \stackrel{\text{def}}{=} \{y_0, y_1, \dots, y_{m-1}\}$ which are also independent identically distributed (iid) and which we may consider as realizations of a Bernoulli random variable \mathbf{y} whose mean is their sample average. We have thus described a noisy channel with input \mathbf{x} and with output \mathbf{y} . In Supplementaries A and B we construct such channels and show that c may be viewed as the correlation between their inputs and outputs.

The *Fisher information* $\mathcal{I}_B(\theta)$ represents the maximum information about θ that Bob may have received by way of the above protocol. If $\mathcal{I}_B(\theta) = \infty$ then Bob knows θ ‘on the nose’, while if $\mathcal{I}_B(\theta) = 0$ then Bob has received no information about θ at all.

Consequence of statistical no-signaling: Given the above setting, if \mathbf{x} and \mathbf{y} are independent random variables then $\mathcal{I}_B(\theta) < \infty$.

When the number of samples is finite, the Fisher information for a disconnected channel obviously vanishes. But when there are infinitely many samples it may happen that $\mathcal{I}_B(\theta) = \infty$. Indeed, we will construct a disconnected channel for which $\mathcal{I}_B(\theta) = \infty$ using superquantum NS-boxes. It is in this way that statistical no-signalling will imply that superquantum NS-boxes are non-physical and therefore that the quantum bound on nonlocality is indeed fundamental.

As we shall see, the only three possible values of $\mathcal{I}_B(\theta)$ in the $m \rightarrow \infty$ limit are 0, 1, and ∞ . When the Fisher information equals zero or one, no information is transmitted about θ . The distinction between these cases is discussed in Supplementary D

To derive the quantum limit on nonlocality $|c| \leq 1/\sqrt{2}$ from statistical no-signaling, we realize a disconnected channel in a specific way as a limiting case of the van Dam protocol [16]. This is the same protocol that was used to test information causality [18].

Alice samples 2^n bits $\mathcal{A} \stackrel{\text{def}}{=} \{\hat{x}_0, \hat{x}_1, \dots, \hat{x}_{2^n-1}\}$ from her ± 1 -valued Bernoulli(θ) random variable \mathbf{x} which she converts into 0/1-valued bits, $\{x_0, x_1, \dots, x_{2^n-1}\}$, such that $\hat{x}_i = (-1)^{x_i+1}$. She then combines these using her NS-boxes, a pair at a time, into one ‘very special’ bit $x^{(n)}$ which she transmits to Bob through what we will for now assume is a perfect channel. Bob randomly chooses an index $0 \leq i \leq 2^n - 1$ (Alice does not know in advance which i he will choose), and makes his best guess y_i (respectively, $\hat{y}_i \stackrel{\text{def}}{=} (-1)^{y_i+1}$) for Alice’s bit x_i (respectively, \hat{x}_i) using $x^{(n)}$ and his NS-boxes. The correlations between Alice’s boxes and Bob’s boxes are governed by the Bell-CHSH correlation $c \in [-1, 1]$. The process described above is called *random access coding* or *oblivious transfer*, and it defines a channel from \mathbf{x} to \mathbf{y} (see Supplementary B and Figure 1). Assume first that $|c| < 1$. A short calculation in Supplementaries B and D will reveal the following properties in the $n \rightarrow \infty$ limit:

- Random variables \mathbf{x} and \mathbf{y} are independent.
- As for the Fisher information:

$$(2) \quad \mathcal{I}_B(\theta) = \lim_{n \rightarrow \infty} \frac{(2c^2)^n}{1 - c^{2n}\theta^2} = \begin{cases} \infty, & 2c^2 > 1 \text{ (signaling)} \\ 1, & 2c^2 = 1 \text{ (randomness)} \\ 0, & 2c^2 < 1 \text{ (no-signaling)} \end{cases}$$

Statistical no-signaling rules out the first case, from which we deduce that $2c^2 \leq 1$, that is the quantum limit on nonlocality (1).

The conceptual explanation for why the channel becomes disconnected as $n \rightarrow \infty$ is that the only information which passes from Alice to Bob in the van Dam protocol is $x^{(n)}$ which is a communication bottleneck. Alice’s information about θ is contained in her samples $x_0, x_1, \dots, x_{2^n-1}$ which are combined with one another and with random noise from the boxes to become $x^{(n)}$. Conversely, Bob’s estimates $y_0, y_1, \dots, y_{2^n-1}$ are also all recovered from $x^{(n)}$ together with noise introduced by his boxes. But $x^{(n)}$ contains less and less information about θ as n grows to infinity and as more boxes are used, and $x^{(n)}$ contains no information at all about θ in the $n \rightarrow \infty$ limit. This disconnects the channel from \mathbf{x} to \mathbf{y} . See Figure 2.

The $|c| = 1$ case (PR-boxes) requires special consideration. The nonlocal correlation c is independent of the characteristics of the classical channel, so we choose the correlation of the classical channel from Alice to Bob in the case of 2^n samples to be $(c')^n$ for some $1/\sqrt{2} < c' < 1$. This disconnects the classical channel between \mathbf{x} to \mathbf{y} , as we show in Supplementary C, while maintaining $\mathcal{I}_B(\theta) = \infty$. This contradicts statistical no-signaling as required.

Our approach differs from others in that we use Fisher information as opposed to Shannon information. As a result, Alice’s bits were interpreted as samples of a random variable whose mean encodes a message. The utility of Fisher information as a measure of the quantity of Bob’s information about θ stems from the

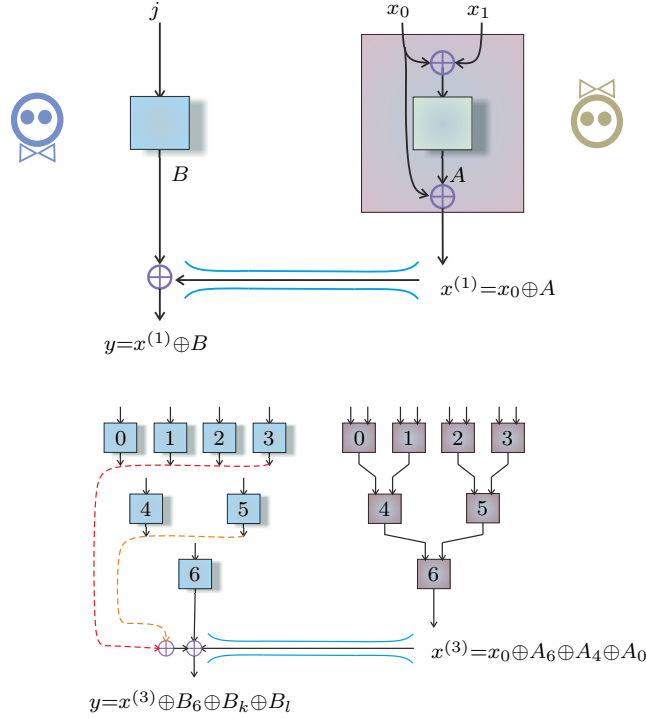


FIGURE 1. Distributed oblivious transfer (van Dam) protocol[16]. Its basic building block is on the left, where Alice inserts $x_0 \oplus x_1$ into her box, receives A , and sends $x_0 \oplus A$ to Bob. Bob decides that he wants to know the value of x_j , and he feeds j into his box, which outputs B . Bob's estimate of x_i is then $x^{(1)} \oplus B$. When there are multiple boxes, Alice concatenates (the process is called *wiring*). For example, with seven boxes, Alice begins with a collection of bits x_0, x_1, \dots, x_7 , and she inputs $x_{2j} \oplus x_{2j+1}$ into box j , where $j = 0, 1, 2, 3$, receiving A_0, A_1, A_2, A_3 correspondingly. The bits fed into the next level of boxes become $x_j^{(1)} \stackrel{\text{def}}{=} x_{2j} \oplus A_j$ with $j = 0, 1, 2, 3$. The final output $x^{(3)}$ is sent to Bob. Bob encodes the address of the bit he wants as the binary number $i_3i_2i_1$ — for example, if he wants x_2 , then he sets $i_3 = 0$, $i_2 = 1$, and $i_1 = 0$ because 10 is 2 in binary. This binary encoding describes a path in his binary tree from a root to a branch, where 0 means ‘go left’ and 1 means ‘go right’. Bob inserts i_3 into the lowermost box to obtain B_6 . Setting $k \stackrel{\text{def}}{=} 5 - (1 - i_3)$, he then inserts i_2 into box k to obtain B_k . Finally, setting $l \stackrel{\text{def}}{=} k - (3 - i_3) - (1 - i_2)$, Bob inserts i_1 into box l to obtain B_l . His final estimate for x_i is $y_i = x^{(3)} \oplus B_6 \oplus B_k \oplus B_l$. Further details are given in Supplementary B.

Cramér–Rao Lower Bound which asserts that $\sigma_B^2 \stackrel{\text{def}}{=} 1/\mathcal{I}_B(\theta)$ is the lowest *uncertainty* about the value of θ in terms of error variance that Bob could hope to achieve with an unbiased estimator.

The Central Limit Theorem (CLT) provides a further interpretation of statistical no-signaling. Let $\bar{\theta}$ be Bob's best decoder of θ based on \mathcal{B} , that is the *maximum likelihood estimator*. In Supplementary E it is shown that as n approaches infinity:

$$(3) \quad \sqrt{\frac{(2c^2)^n}{1 - c^{2n}\theta^2}} (\bar{\theta} - \theta) \xrightarrow{d} \nu \sim \mathcal{N}(0, 1) ,$$

where d stands in for convergence in distribution. The rightmost term denotes a random variable whose distribution is Gaussian centered at 0 with variance 1. We may think of (3) as a form of the CLT in which the *number of samples* has been replaced by the Fisher information. Explicitly, for $c = 1$ and for $\theta = 0$ we recover the usual CLT. Thinking of the Fisher information as the *effective number of samples* that Bob receives, if $2c^2 \leq 1$ then (3) appears as a retarded or degenerate CLT in which the effective number of samples does not increase as $n \rightarrow \infty$. Thus Bob's ability to estimate θ using $\bar{\theta}$ decreases in the $n \rightarrow \infty$

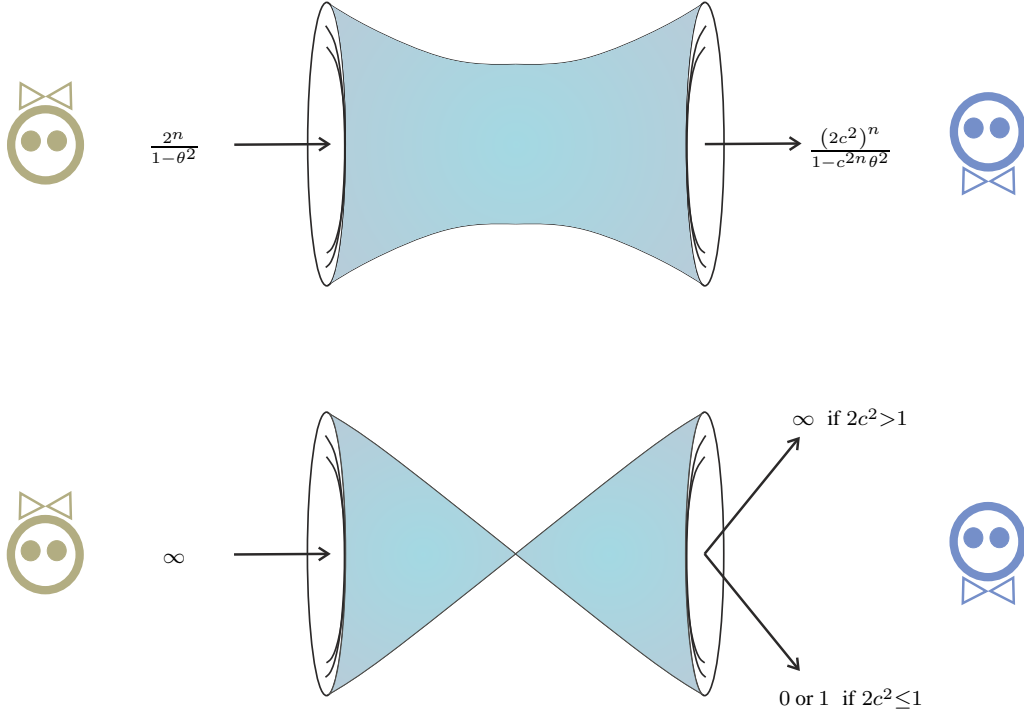


FIGURE 2. The statistical no-signaling condition. The van Dam protocol defines an underlying channel which becomes disconnected in the $n \rightarrow \infty$ limit. The upper illustration shows this channel and the amount of Fisher information about θ at its input and at its output. When the number of nonlocal resources increases unboundedly, the two ends of the channel become disconnected as illustrated by a vanishing bottleneck in the lower figure. Statistical no-signaling dictates that in this case no information can pass through, which occurs if and only if $2c^2 \leq 1$. The case of $2c^2 > 1$ leads to a physically unreasonable limit where Bob can fully read off the value of Alice's θ through a disconnected channel.

limit, which is what we would expect because less information about θ is passing through the channel. Despite the number of samples growing, the effective number of samples does not increase.

Conclusions

We have formulated a *statistical no-signaling* principle which dictates that no information can pass through a disconnected channel. Applied to an infinite limit of the van Dam protocol, this principle is equivalent to the quantum bound on nonlocality. We may view this fact as an example of asymptotic theory in statistics, in which an asymptotic limit allows us to discern statistical properties that are unavailable for a finite number of samples.

Statistical no-signaling is different from the notion of *no-signaling* in the sense of non-signaling theories (NS-boxes). No-signaling pertains to a single pair of boxes, whereas statistical no-signaling is used as a condition on the limit of an iterative construction involving infinitely many boxes. Taking statistical no-signaling instead of what is traditionally called ‘no-signaling’ as our no-signaling condition, we recover the idea that quantum mechanics is indeed the most general nonlocal non-signaling theory.

Shimony [10, 11] and Aharonov [12] independently suggested that a quantum theory may perhaps be based on two axioms, nonlocality and relativistic causality (no-signaling). Aharonov (unpublished) also observed that these two axioms, which seem to contradict one another, can be reconciled using *uncertainty*. This idea was virtually abandoned for many years following the discovery of superquantum theories which satisfy both axioms. But statistical no-signaling reveals a sense in which this original idea holds true. When the number of nonlocal resources increases to infinity, stronger-than-quantum nonlocal theories fail to satisfy statistical no-signaling. These superquantum theories approach a signaling limit where Bob can

recover Alice’s message with complete certainty even though the channel between Alice and Bob is disconnected. Quantum nonlocality obeys statistical no-signaling and thus permits only bounded uncertainty (pure randomness), $\sigma_B^2 \rightarrow 1$, or complete uncertainty, $\sigma_B^2 \rightarrow \infty$, in the limit.

The statistical no-signaling condition is stronger than previously identified principle of information causality [18]. Violation of statistical no-signaling implies violation of information causality whereas the converse implication is false. This is evident in the derivation of information causality in that paper, where the expression of the Fisher information in (2) in the $\theta = 0$ case appears as Equation (23) therein.

References

- [1] Bell, J.S. On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200 (1964).
- [2] Popescu, S., Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys* **24**, 379–385 (1994).
- [3] Popescu, S. Nonlocality beyond quantum mechanics. *Nature Phys.* **10**, 264–270 (2014).
- [4] Oas, G., de Barros, J.A. A Survey of Physical Principles Attempting to Define Quantum Mechanics. [arXiv:1506.05515](#)
- [5] Poh, H.S., Joshi, S.K., Ceré, A., Cabello, A., Kurtsiefer, C. Approaching Tsirelson’s bound in a photon pair experiment. [arXiv:1506.01865](#)
- [6] Clauser, J., Horne, M., Shimony, A., Holt, R. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
- [7] Hensen, B. et al. Experimental loophole-free violation of a Bell inequality using entangled electron spins separated by 1.3 km. [arXiv:1508.05949](#)
- [8] Cirel’son, B. S. Quantum generalizations of Bell’s inequality. *Lett. Math. Phys.* **4**, 93–100 (1980).
- [9] Seife, C. Do deeper principles underlie quantum uncertainty and nonlocality? *Science* **309**, 98 (2005).
- [10] Shimony, A. in *Proceedings of the International Symposium on Foundations of Quantum Mechanics in the Light of New Technology* (S.Kamefuchi et al., Eds.) in 225 (Phys.Soc.Japan, 1984).
- [11] Shimony, A. in *Quantum Concepts of Space and Time*, (Penrose, R., Isham, C., Eds.) in 182 (Oxford University Press, 1986).
- [12] Aharonov, Y., Rohrlich, D. *Quantum Paradoxes: Quantum Theory for the Perplexed* Chapter 6.4, (Wiley-VCH, 2005).
- [13] Barrett, J. et al. Non-local correlations as an information theoretic resource. *Phys. Rev. A* **71**, 022101 (2005). [arXiv:quant-ph/0404097](#)
- [14] Wolf, M., Garcia, D.P., Fernandez, C. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Phys. Rev. Lett.* **103**, 230402 (2009). [arXiv:0905.2998](#)
- [15] Oppenheim, J., Wehner, S. The uncertainty principle determines the non-locality of quantum mechanics. *Science* **330**, 1072–1074 (2010). [arXiv:1004.2507](#)
- [16] van Dam, W. Implausible consequences of superstrong nonlocality. *Nat. Comput.* **12**(1), 9–12 (2013). [arXiv:quant-ph/0501159](#)
- [17] Linden, N., Popescu, S., Short, A.J., Winter, A. Quantum nonlocality and beyond: limits from nonlocal computation. *Phys. Rev. Lett.* **99**, 180502 (2007).
- [18] Pawłowski, M. et al. Information causality as a physical principle. *Nature* **461**, 1101–1104 (2009).
- [19] Rohrlich, D. PR-box correlations have no classical limit. in *Quantum Theory: A Two-Time Success Story* in 205–211 (Springer, 2014).
- [20] Navascués, M., Wunderlich, H. A glance beyond the quantum model. *Proc. Royal Society A.* **466**, 881–890 (2010). [arXiv:0907.0372](#)
- [21] Navascués, M., Guryanova, Y., Hoban, M.J., Acín, A., Almost quantum correlations. *Nat. Commun.* **6**, 6288, (2015). [arXiv:1403.4621](#)

Acknowledgements.: The authors thank D. Rohrlich for useful discussions.

Supplementary A. The bipartite Bell experiment as a noisy symmetric channel

In this section we recall the definition of the Bell–CHSH correlation c and we formulate the Bell–CHSH inequality, establishing notation. We then exhibit c as the correlation of a symmetric binary channel.

A.1. The Bell–CHSH inequality. Let us recall the classical bipartite Bell experiment.

Alice and Bob each hold one half of an EPR pair such as a singlet state of spin- $\frac{1}{2}$ particles. They each possess two different measuring instruments which we unimaginatively call ‘instrument zero’ and ‘instrument one’. Alice measures her particle using one of the instruments, and Bob does the same. Let a be the index of the instrument used by Alice and let \hat{A} be its reading. Similarly, let b and \hat{B} denote the index of an instrument chosen by Bob and its reading. In the language of probability, \hat{A} and \hat{B} are ± 1 -valued Bernoulli random variables. The choices of measuring instrument, a and b , may be either parameters or 0/1-valued Bernoulli random variables.

Repeating the experiment for many different EPR pairs, Alice and Bob may compute the correlations of their readings \hat{A} and \hat{B} for any given pair of indices a and b . Formally, they compute, $E[\hat{A}\hat{B} \mid a, b]$, the expectation of $\hat{A}\hat{B}$ conditioned on their choice of a particular pair of measuring instruments a and b . We now define the *Bell–CHSH correlation* c by the formula:

$$(4) \quad 4c \stackrel{\text{def}}{=} E[\hat{A}\hat{B} \mid 0, 0] + E[\hat{A}\hat{B} \mid 0, 1] + E[\hat{A}\hat{B} \mid 1, 0] - E[\hat{A}\hat{B} \mid 1, 1] .$$

In any theory in which both Alice and Bob’s choices, and the readings of their measuring devices, are *local*, the Bell–CHSH inequality [6] holds:

$$(5) \quad |c| \leq \frac{1}{2} .$$

Locality means that Alice’s readings may only be affected by her own choices (or perhaps by any other hidden variables locally at her site), and similarly for Bob’s readings. Quantum mechanically, Alice and Bob may violate (5) and hence Quantum Mechanics is *nonlocal*.

A.2. The Bell–CHSH correlation c as a channel correlation. Non-signaling (NS)–boxes provide an abstraction and an extension of the Bell–CHSH experiment. This time, Alice and Bob each owns a box. Each such box may be thought of as a complete laboratory containing two measuring devices. Either participants inserts their choice of measuring device into their box. The box output is the respective reading of the chosen measuring device.

Alice and Bob share a pair of NS–boxes whose inputs are a and b and whose outputs are Bernoulli random variables A_a and B_b . Assume now that a, b, A_a , and B_b are all 0/1-valued.

We will show that the Bell–CHSH parameter (4) represents the correlation of a symmetric binary channel whose input is the Bernoulli random variable $\mathbf{x} \stackrel{\text{def}}{=} \widehat{ab}$ and whose output is the Bernoulli random variable $\mathbf{y} \stackrel{\text{def}}{=} \widehat{A_a \oplus B_b}$, where \hat{f} denotes $(-1)^f$.

Let $i \in \{0, 1\}$. Define *channel correlations* c_i as follows:

$$(6) \quad c_i \stackrel{\text{def}}{=} E[\mathbf{xy} \mid \mathbf{x} = \hat{i}] = P(\mathbf{y} = \hat{i} \mid \mathbf{x} = \hat{i}) - P(\mathbf{y} \neq \hat{i} \mid \mathbf{x} = \hat{i}) = 2P(\mathbf{y} = \hat{i} \mid \mathbf{x} = \hat{i}) - 1 .$$

With respect to a particular choice of measuring devices a and b , (6) becomes:

$$(7) \quad c_i(a, b) = E[\widehat{A_a \oplus B_b ab} \mid a, b, \hat{ab} = \hat{i}] = 2P(A_a \oplus B_b = ab \mid a, b, ab = i) - 1 .$$

Pulling the condition $\hat{i} = \hat{ab} = (-1)^{ab}$ out of (7) and using $\widehat{A_a \oplus B_b} = \hat{A}_a \hat{B}_b$, we obtain:

$$(8) \quad c_{\hat{ab}}(a, b) = E[\hat{A}_a \hat{B}_b \hat{ab} \mid a, b] = (-1)^{ab} E[\hat{A}_a \hat{B}_b \mid a, b] .$$

Assume that the channel is symmetric, *i.e.* that $c = c_{ab}(a, b)$, $\forall a, b$. From (7) and (8) we may rewrite the Bell-CHSH correlation (4) as:

$$\begin{aligned}
 (9) \quad c &= \frac{1}{4} (c_1(0, 0) + c_1(0, 1) + c_1(1, 0) + c_{-1}(1, 1)) \\
 &= \frac{1}{4} \left\{ E \left[\hat{A}_a \hat{B}_b \mid 0, 0 \right] + E \left[\hat{A}_a \hat{B}_b \mid 0, 1 \right] + E \left[\hat{A}_a \hat{B}_b \mid 1, 0 \right] - E \left[\hat{A}_a \hat{B}_b \mid 1, 1 \right] \right\} \\
 &= 2P(A_a \oplus B_b = ab \mid a, b) - 1 .
 \end{aligned}$$

The last equality above follows from the channel symmetry:

$$\begin{aligned}
 (10) \quad c &= 2P \left(A_a \oplus B_b = ab \mid a, b, ab = 0 \right) - 1 = \\
 &2P \left(A_a \oplus B_b = ab \mid a, b, ab = 1 \right) - 1 = 2P \left(A_a \oplus B_b = ab \mid a, b \right) - 1 .
 \end{aligned}$$

Equation (9) is our promised interpretation of the Bell-CHSH correlation as a correlation of a noisy symmetric binary channel.

Supplementary B. The van Dam protocol as a noisy symmetric channel

In this section we recall the construction of the van-Dam protocol [18, 16]. We then reinterpret this protocol as underlying a noisy symmetric binary channel, as a special case of the construction of Section A. We compute its correlations, and establish the effect of noise on its classical component.

B.1. The van Dam protocol. The van Dam protocol realizes an *oblivious transfer protocol* by means of a classical channel and a number of NS-boxes. Each of Alice's boxes has a corresponding box on Bob's side, and different pairs of boxes are statistically independent. Suppose that Alice has in her possession the bits x_0, \dots, x_{m-1} where $m = 2^n$, $n \geq 1$. Bob wishes to know the value of one of her bits. He may do so by specifying the address of the bit whose value he wishes to know via its binary address $i = 1_{n-1}i_{i-2} \dots i_0$. For example, if $n = 2$ then Bob may specify which of the bits x_0 to x_3 he wants by specifying a binary address, 00, 01, 10, or 11. Alice bits and Bob addresses are encoded into the inputs of $2^n - 1$ NS-boxes following a particular protocol which is described next.

Alice uses outputs of boxes and choices of measuring devices to determine choices of measuring devices for other boxes. Such a procedure is called *wiring*. The wiring of boxes on Alice side admits a recursive description which we now give. Let $A_a^{k,j}$ denote the output of the j th box on the k th level on Alice side. Let also:

$$(11) \quad f^{k,j}(q_1, q_2) \stackrel{\text{def}}{=} q_1 \oplus A_{q_1 \oplus q_2}^{k,j} .$$

Suppose that Alice wishes to encode $m = 4$ bits with her boxes. To do so, she first picks two boxes and computes:

$$(12) \quad x_1^{(1)} \stackrel{\text{def}}{=} f^{1,1}(x_0, x_1), \quad x_2^{(1)} \stackrel{\text{def}}{=} f^{1,2}(x_2, x_3) .$$

This forms the first level in her construction. The second level then follows:

$$(13) \quad x^{(2)} \stackrel{\text{def}}{=} f^{2,1}(x_1^{(1)}, x_2^{(1)}) .$$

In this example there are only two levels and so $x^{(2)}$ is the bit which Alice transmits to Bob through the classical channel. In case where $m = 2^n$ there will be n levels and thus $x^{(n)}$ is the bit Bob will receive from Alice.

Unbeknownst to Alice, Bob now decides which bit x_i he would like to know the value of. He takes its binary address $i = i_{n-1}i_{i-2} \dots i_0$, and inserts i_{k-1} into all of his boxes whose counterparts are on the k level on Alice's side. He then uses the values $B_{i_{k-1}}^{k,j}$ that he obtains, together with the bit $x^{(n)}$ he received from Alice, to construct the decoding function:

$$(14) \quad y_i \stackrel{\text{def}}{=} x^{(n)} \oplus B_{i_0}^{1,j_1} \oplus B_{i_1}^{2,j_2} \oplus \dots \oplus B_{i_{n-1}}^{n,j_n} .$$

The values j_1, \dots, j_n (which boxes Bob uses) are determined by the binary address $i = i_{n-1}i_{n-2} \dots i_0$ via the recursive formula $j_{l-1} = 2j_l - 1 + i_{l-1}$ for $l = 1, 2, \dots, n-1$ starting from $j_n = 1$.

The probability that Bob will decode the correct value of the bit he desires is governed by the NS-box correlation c . For the simplest case of $m = 2$ where Alice and Bob share a single pair of boxes, note that

$$(15) \quad 2P\left(y_i = x_{i_1} \mid x_{i_1}\right) - 1 = 2P\left(f^{1,1}(x_0, x_1) \oplus B_{i_1}^{1,1} = x_{i_1} \mid x_{i_1}\right) - 1 \\ = 2P\left(x_0 \oplus A_{x_0 \oplus x_1}^{1,1} \oplus B_{i_1}^{1,1} = x_{i_1} \mid x_{i_1}\right) - 1 .$$

As $x_{i_1} = x_0 \oplus i_1(x_0 \oplus x_1)$, this equals:

$$(16) \quad 2P\left(x_0 \oplus A_{x_0 \oplus x_1}^{1,1} \oplus B_{i_1}^{1,1} = x_0 \oplus i_1(x_0 \oplus x_1) \mid x_{i_1}\right) - 1 \\ = 2P\left(A_{x_0 \oplus x_1}^{1,1} \oplus B_{i_1}^{1,1} = ab \mid a = x_0 \oplus x_1, b = i_1, x_{i_1}\right) - 1 = c ,$$

which follows from (9).

In general, decoding any bit out of 2^n possible bits involves using n pairs of NS boxes. Noting that an even number of errors, $A \oplus B \neq ab$, will always cancel out in such a construction, leads to [18]:

$$(17) \quad c^n = 2P\left(y_i = x_i \mid x_i\right) - 1 .$$

We illustrate in the case that $n = 2$:

$$(18) \quad P\left(A_{a_1} \oplus B_{b_1} \oplus A_{a_2} \oplus B_{b_2} = a_1b_1 \oplus a_2b_2 \mid a_{1,2}, b_{1,2}, a_1b_1 \oplus a_2b_2\right) = \\ P\left(A_{a_1} \oplus B_{b_1} = a_1b_1 \mid a_1, b_1\right) P\left(A_{a_2} \oplus B_{b_2} = a_2b_2 \mid a_2, b_2\right) + \\ P\left(A_{a_1} \oplus B_{b_1} \neq a_1b_1 \mid a_1, b_1\right) P\left(A_{a_2} \oplus B_{b_2} \neq a_2b_2 \mid a_2, b_2\right) = \\ \frac{1}{2}(1+c) \cdot \frac{1}{2}(1+c) + \frac{1}{2}(1-c) \cdot \frac{1}{2}(1-c) = \frac{1}{2}(1+c^2) .$$

B.2. van Dam protocol as a symmetric channel. Assume now that instead of a string of bits, Alice has in her possession an information source that is a ± 1 -valued Bernoulli random variable \mathbf{x} whose mean is θ . Alice generates m iid samples, $\hat{x}_0, \dots, \hat{x}_{m-1}$ from \mathbf{x} and converts them into her 0/1-valued bits, x_0, x_1, \dots, x_{m-1} by mapping 0 to -1 and 1 to 1. As in (18), the van Dam protocol has a *memoriless* property:

$$(19) \quad P\left(y_i = x_i \mid x_0, x_1, \dots, x_{m-1}\right) = P\left(y_i = x_i \mid x_i\right) .$$

From this it follows that if Alice's inputs x_0, x_1, \dots, x_{m-1} are iid then Bob's outputs y_0, y_1, \dots, y_{m-1} are also iid. Therefore the set $\hat{y}_i \stackrel{\text{def}}{=} (-1)^{y_i}$ determine a Bernoulli random variable \mathbf{y} . In this way, the van Dam protocol may be viewed as a symmetric binary channel whose input is \mathbf{x} and whose output is \mathbf{y} . By (17) the channel correlation is

$$(20) \quad E[\mathbf{xy} \mid \mathbf{x} = \hat{x}_i] = 2P\left(\mathbf{y} = \hat{x}_i \mid \mathbf{x} = \hat{x}_i\right) - 1 = 2P\left(y_i = x_i \mid x_i\right) - 1 = c^n .$$

B.3. Noisy classical channel in the van Dam protocol. The preceding discussion of the van Dam protocol assumed a perfect classical channel between Alice and Bob. We now relax this assumption. Let $(c')^n$ be the correlation underlying the classical channel, where $|c'| \leq 1$. Such a channel can be realized by concatenating n copies of a noisy symmetric channel whose correlation is c' . This correlation depends on n , and Alice may construct it as part of the protocol based on her knowledge of n .

Note first that:

$$(21) \quad P(\mathbf{z} = i \mid \mathbf{x} = i) = P(\mathbf{z} = i \mid \mathbf{y} = i) P(\mathbf{y} = i \mid \mathbf{x} = i) + P(\mathbf{z} = i \mid \mathbf{y} \neq i) P(\mathbf{y} \neq i \mid \mathbf{x} = i) = P(\mathbf{z} = i \mid \mathbf{y} = i) P(\mathbf{y} = i \mid \mathbf{x} = i) + P(\mathbf{z} \neq i \mid \mathbf{y} = i) P(\mathbf{y} \neq i \mid \mathbf{x} = i) .$$

Let \mathbf{y} and \mathbf{z} be the input and output of a symmetric classical channel. By (6) we may write:

$$(22) \quad (c')^n = E[\mathbf{y}\mathbf{z}] = 2P(\mathbf{z} = i \mid \mathbf{y} = i) - 1 ,$$

and similarly we may rewrite (20) as:

$$(23) \quad c^n = E[\mathbf{x}\mathbf{y}] = 2P(\mathbf{y} = i \mid \mathbf{x} = i) - 1 .$$

Substituting (22) and (23) into (21) gives us that for the van Dam protocol with a noisy classical channel:

$$(24) \quad P(\mathbf{z} = i \mid \mathbf{x} = i) = [1 + (cc')^n] / 2 .$$

From this we see that $(cc')^n = E[\mathbf{x}\mathbf{z}]$ is the correlation of the symmetric binary channel defined by the van Dam protocol in the case of a classical channel with correlation $(c')^n$ and a Bell-CHSH correlation c .

Supplementary C. The van Dam channel disconnects in the $n \rightarrow \infty$ limit

If $|c| < 1$ or $|c'| < 1$ then it follows that:

$$(25) \quad E[\mathbf{x}\mathbf{z}] = 2P(\mathbf{z} = i \mid \mathbf{x} = i) - 1 = (cc')^n \xrightarrow{n \rightarrow \infty} 0 .$$

Therefore, in the $n \rightarrow \infty$ limit:

$$(26) \quad P(\mathbf{z} = i \mid \mathbf{x} = i) = 1/2 .$$

But also:

$$(27) \quad P(\mathbf{z} = i) = P(\mathbf{z} = i \mid \mathbf{x} = i) P(\mathbf{x} = i) + P(\mathbf{z} = i \mid \mathbf{x} \neq i) P(\mathbf{x} \neq i) = \frac{1}{2}(P(\mathbf{x} = i) + P(\mathbf{x} \neq i)) = \frac{1}{2} .$$

Combining (26) with (27) gives us that:

$$(28) \quad P(\mathbf{z} \mid \mathbf{x}) \xrightarrow{n \rightarrow \infty} P(\mathbf{z}) .$$

Thus \mathbf{x} and \mathbf{z} are statistically independent in the $n \rightarrow \infty$ limit.

Supplementary D. Transmission of Fisher information through binary channels

In this section we compute the Fisher information of samples of a Bernoulli random variable sent through a binary channel about the mean of the information source. Applying this to the symmetric binary channel that underlies the van Dam protocol, we obtain (2) in the main text. We also discuss the case that the quantum limit on nonlocality is attained, *i.e.* that $|cc'| = 1/\sqrt{2}$.

D.1. Fisher information for a binary channel. Consider a *binary channel* whose input is a ± 1 -valued Bernoulli random variable \mathbf{x} and whose output is another ± 1 -valued Bernoulli random variable \mathbf{y} . The channel correlations are defined by means of (6). If the channel is symmetric then

$$(29) \quad P(\mathbf{y} = \mathbf{x}) = P(\mathbf{y} = -1 \mid \mathbf{x} = -1) = P(\mathbf{y} = 1 \mid \mathbf{x} = 1) ,$$

from which it follows that $c_{-1} = c_1 = E[\mathbf{x}\mathbf{y}]$.

We shall assume a prior distribution for \mathbf{x} given by:

$$(30) \quad P(\mathbf{x} = -1 \mid \theta) = \frac{1}{2}(1 + \theta) ,$$

with parameter $\theta \in [-1, 1]$. Using this we may write

$$(31) \quad \begin{aligned} P(\mathbf{y} = -1 \mid \theta) &= P(\mathbf{y} = -1 \mid \mathbf{x} = -1) P(\mathbf{x} = -1 \mid \theta) + P(\mathbf{y} = -1 \mid \mathbf{x} = 1) P(\mathbf{x} = 1 \mid \theta) = \\ &= \frac{1}{2} \left[1 + \frac{1}{2}(c_{-1} - c_1) + \frac{1}{2}(c_{-1} + c_1)\theta \right] . \end{aligned}$$

Alice sends m iid random samples $\mathcal{X} \stackrel{\text{def}}{=} \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ through the channel. Denote the set of respective outputs $\mathcal{Y} \stackrel{\text{def}}{=} \{\mathbf{y}_1, \dots, \mathbf{y}_m\}$. The *likelihood* of θ given the set \mathcal{Y} is given by the expression:

$$(32) \quad P(\mathcal{Y} \mid \theta) = \left[P(\mathbf{y} = -1 \mid \theta) \right]^{\sum_{i=1}^m \mathbf{1}_{\{\mathbf{y}_i = -1\}}} \left[P(\mathbf{y} = 1 \mid \theta) \right]^{\sum_{i=1}^m \mathbf{1}_{\{\mathbf{y}_i = 1\}}} ,$$

where the *indicator* random variable of a random event A is given as:

$$(33) \quad \mathbf{1}_A \stackrel{\text{def}}{=} \begin{cases} 1, & A \text{ occurred;} \\ 0, & \text{otherwise.} \end{cases}$$

According to (32) the log-likelihood is given by the expression:

$$(34) \quad \mathcal{L}(\theta) \stackrel{\text{def}}{=} \log P(\mathcal{Y} \mid \theta) = \left[\sum_{i=1}^m \mathbf{1}_{\{\mathbf{y}_i = -1\}} \right] \log P(\mathbf{y} = -1 \mid \theta) + \left[\sum_{i=1}^m \mathbf{1}_{\{\mathbf{y}_i = 1\}} \right] \log P(\mathbf{y} = 1 \mid \theta) .$$

The *Fisher information* about θ contained in the set \mathcal{Y} is defined as:

$$(35) \quad \mathcal{I}_{\mathcal{Y}}(\theta) \stackrel{\text{def}}{=} E \left[\left(\frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right)^2 \right] = -E \left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2} \right] .$$

Note that:

$$(36) \quad E \left[\sum_{i=1}^m \mathbf{1}_{\{\mathbf{y}_i = s\}} \right] = \sum_{i=1}^m E [\mathbf{1}_{\{\mathbf{y}_i = s\}}] = mP(\mathbf{y} = s \mid \theta), \quad s = -1, 1 .$$

Using this, (35) reads:

$$(37) \quad \mathcal{I}_{\mathcal{Y}}(\theta) = \frac{m \left[\frac{1}{2}(c_{-1} + c_1) \right]^2}{1 - \left[\frac{1}{2}(1 + \theta)c_{-1} - \frac{1}{2}(1 - \theta)c_1 \right]^2} .$$

For a symmetric binary channel, with $c = c_{-1} = c_1$, Equation (37) simplifies to:

$$(38) \quad \mathcal{I}_{\mathcal{Y}}(\theta) = \frac{mc^2}{1 - c^2\theta^2} .$$

Note that the minimum of $\mathcal{I}_{\mathcal{Y}}(\theta)$ is obtained for $\theta = 0$ in which case $P(\mathbf{x} \mid \theta) = 1/2$ and $\mathcal{I}_{\mathcal{Y}}(0) = mc^2$.

D.2. Fisher information in the van Dam protocol. Alice begins with $m = 2^n$ iid samples. As shown in Section B, the van Dam protocol defines a channel from Alice to Bob whose correlation is $(cc')^n$ where c is the Bell–CHSH correlation and c' is the correlation of the classical channel used by Alice in her construction. The channel input and output are the random variables \mathbf{x} and \mathbf{z} . The maximum amount of Fisher information that Bob may receive about θ is attained after he has asked for all of Alice bits, *i.e.* after repeating the protocol 2^n times where in each run Bob inputs the indices of his newly requested bit into his boxes. Let \mathcal{B} be the set of outputs on Bob's end. According to (38):

$$(39) \quad \mathcal{I}_{\mathcal{B}}(\theta) = \frac{[2(cc')^2]^n}{1 - (cc')^{2n}\theta^2} .$$

This simplifies to (2) when $|c'| = 1$.

D.3. Interpretation of the case in which the quantum limit on nonlocality is attained. Fisher information about θ is a function of θ . We see in (39) that this function is identically 1 for $2(cc')^2 = 1$ in the $n \rightarrow \infty$ limit. One is also the minimum amount of Fisher information contained in a single bit on Alice end, namely, when $n = 0$ and $\theta = 0$. When the quantum limit on nonlocality is attained, Bob could receive the same amount of information about θ as he could attain by the van Dam protocol by just tossing a fair coin. We thus consider this $\mathcal{I}_{\mathcal{B}}(\theta) \equiv 1$ case to be an instance of no-signaling in which Bob's decoded bits carry no information about the actual value of θ .

Supplementary E. Statistical no-signaling and the Central Limit Theorem

Bob may estimate the quantity $c^n\theta$ from his decoded iid samples by computing the sample mean:

$$(40) \quad c^n\bar{\theta} \stackrel{\text{def}}{=} \frac{1}{2^n} \sum_{i=0}^{2^n-1} \hat{y}_i ,$$

which, by the strong law of large numbers, is unbiased and converges almost surely to $c^n\theta$. This is also a maximum likelihood estimator as its variance attains the Cramér–Rao bound. In particular,

$$(41) \quad \text{Var}(c^n\bar{\theta}) = c^{2n} \text{Var}(\bar{\theta}) = \frac{\text{Var}(\mathbf{y})}{2^n} = \frac{1 - c^{2n}\theta^2}{2^n} \longrightarrow \text{Var}(\bar{\theta}) = 1/\mathcal{I}_{\mathcal{B}}(\theta) .$$

The Central Limit Theorem governs the convergence of $c^n\bar{\theta}$ to $c^n\theta$ as $n \rightarrow \infty$:

$$(42) \quad \sqrt{\frac{2^n}{\text{Var}(\mathbf{y})}} (c^n\bar{\theta} - c^n\theta) \xrightarrow{d} \nu \sim \mathcal{N}(0, 1) ,$$

where d means *convergence in distribution*. Thus:

$$(43) \quad \sqrt{\frac{(2c^2)^n}{1 - c^{2n}\theta^2}} (\bar{\theta} - \theta) \xrightarrow{d} \nu \sim \mathcal{N}(0, 1) .$$